



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle

Citation for published version:

European Cattle Genetic Diversity Consortium, Utsunomiya, YT, Bomba, L, Lucente, G, Colli, L, Negrini, R, Lenstra, JA, Erhardt, G, Garcia, JF & Ajmone-Marsan, P 2014, 'Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle', *BMC Genetics*, vol. 15, no. 1, 47. <https://doi.org/10.1186/1471-2156-15-47>

Digital Object Identifier (DOI):

[10.1186/1471-2156-15-47](https://doi.org/10.1186/1471-2156-15-47)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

BMC Genetics

Publisher Rights Statement:

© 2014 Utsunomiya et al.; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



RESEARCH ARTICLE

Open Access

Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle

Yuri Tani Utsunomiya^{1†}, Lorenzo Bomba^{2†}, Giordana Lucente², Licia Colli^{2,3}, Riccardo Negrini², Johannes Arjen Lenstra⁴, Georg Erhardt⁵, José Fernando Garcia^{1,6}, Paolo Ajmone-Marsan^{2,3*} and European Cattle Genetic Diversity Consortium

Abstract

Background: Descendants from the extinct aurochs (*Bos primigenius*), taurine (*Bos taurus*) and zebu cattle (*Bos indicus*) were domesticated 10,000 years ago in Southwestern and Southern Asia, respectively, and colonized the world undergoing complex events of admixture and selection. Molecular data, in particular genome-wide single nucleotide polymorphism (SNP) markers, can complement historic and archaeological records to elucidate these past events. However, SNP ascertainment in cattle has been optimized for taurine breeds, imposing limitations to the study of diversity in zebu cattle. As amplified fragment length polymorphism (AFLP) markers are discovered and genotyped as the samples are assayed, this type of marker is free of ascertainment bias. In order to obtain unbiased assessments of genetic differentiation and structure in taurine and zebu cattle, we analyzed a dataset of 135 AFLP markers in 1,593 samples from 13 zebu and 58 taurine breeds, representing nine continental areas.

Results: We found a geographical pattern of expected heterozygosity in European taurine breeds decreasing with the distance from the domestication centre, arguing against a large-scale introgression from European or African aurochs. Zebu cattle were found to be at least as diverse as taurine cattle. Western African zebu cattle were found to have diverged more from Indian zebu than South American zebu. Model-based clustering and ancestry informative markers analyses suggested that this is due to taurine introgression. Although a large part of South American zebu cattle also descend from taurine cows, we did not detect significant levels of taurine ancestry in these breeds, probably because of systematic backcrossing with zebu bulls. Furthermore, limited zebu introgression was found in Podolian taurine breeds in Italy.

Conclusions: The assessment of cattle diversity reported here contributes an unbiased global view to genetic differentiation and structure of taurine and zebu cattle populations, which is essential for an effective conservation of the bovine genetic resources.

Keywords: Cattle, AFLP, Genetic differentiation, Ascertainment bias

* Correspondence: paolo.ajmone@unicatt.it

†Equal contributors

²Institute of Zootechnics, Università Cattolica del Sacro Cuore, Piacenza, Italy

³BioDNA Biodiversity and Ancient DNA Research Centre, Università Cattolica del Sacro Cuore, Piacenza, Italy

Full list of author information is available at the end of the article

Background

As a consequence of over 10,000 years of domestication, migrations and natural as well as artificial selection, a wide range of phenotypically distinct cattle populations spread around the world. Several research initiatives have combined molecular marker datasets with historic and archaeological records in order to investigate the origin, history, genetic diversity, and differentiation of cattle populations (see Groeneveld *et al.*, 2010 [1] for a review on the topic). The collected evidences suggest that domestic cattle descend from the extinct aurochs (*Bos primigenius*) and are divided into two distinct but interfertile species: the humpless taurine cattle (*Bos taurus*) and the humped indicine or zebu cattle (*Bos indicus*). It is accepted that taurine and zebu cattle have arisen from separate centres of domestication about 8,000 years BC in the Fertile Crescent (modern-day countries of Israel, Jordan, Lebanon, Cyprus and Syria, and parts from Egypt, Turkey, Iraq, Iran and Kuwait) and the Indus valley (current Pakistan), respectively [2,3]. From these regions, cattle have spread throughout Europe, Asia and Africa due to the expansion of agriculture [4,5]. Taurine cattle were imported to the American continent after 1492, mainly from Iberian importations; in the early 20th century, Indian zebu cattle were introduced in Central and South America because of their adaptability to the tropical environment.

Molecular markers have been essential to the investigation of the history and genetic differentiation of domestic cattle. Recent studies applying genome-wide single nucleotide polymorphism (SNP) markers to investigate genetic structure and differentiation in multiple cattle breeds (e.g., [6-9]) resolved hypotheses that were not possible to be tested by using sparse panels of molecular markers. However, markers included in the most widely used SNP panel, the Illumina® BovineSNP50 BeadChip assay (50 k), were discovered in reduced representation libraries from pooled DNA samples of six taurine breeds [10], which leads to biased estimates of genetic structure and differentiation in zebu cattle [7].

As an alternative, amplified fragment length polymorphism (AFLP) markers [11] have been used for almost two decades. Due to their random nature and high reproducibility, they have been enabling ascertainment bias-free analysis of diversity in any species since before the advent of high throughput genotyping and sequencing technologies [12]. AFLP markers are produced by digesting genomic DNA with both a rare cutter and a frequent cutter restriction enzyme, with subsequent ligation of synthetic adapters to the restriction fragments to serve as primer-binding sites, and selective amplification of subsets of the restriction fragments with primers carrying additional nucleotides at their 3' end [13].

Although AFLP markers are highly informative [13-15] and unbiased, there are few examples of the application of

this type of marker in multiple breed, large-scale population differentiation analysis in cattle. Negrini *et al.* [16] used 81 AFLP and 19 microsatellite markers to estimate genetic distances among 51 breeds of cattle, including taurine and zebu cattle, and found that the AFLP panel could differentiate between zebu and taurine cattle better than the panel of microsatellites. Two studies in pigs [17,18] showed the potential of AFLP to survey genetic diversity at the continental scale. Because AFLP polymorphisms are mainly (but not exclusively) based on point mutations, these markers are expected to indicate evolutionary divergence better than microsatellites with variable mutation rates. For instance, a microsatellite-based bovine phylogeny [19] was not in agreement with a phylogeny based on sequence data [20], which was not the case for an AFLP-based phylogeny [21]. Thus, AFLP appears to be a valuable complementary tool for studies of genetic diversity in cattle populations around the world.

Aiming at an unbiased view of genetic structure and differentiation between taurine and zebu cattle breeds from distinct continental areas, we compiled a worldwide multi-breed AFLP dataset. We do not intend to suggest the use of sparse panels of molecular markers over the present portfolio of high-density SNP arrays, or to interrogate their legitimacy for diversity research in cattle. Instead, we intend to propose an unbiased model of cattle differentiation which complements the assessment of genetic distance estimates obtained from molecular markers that are likely to suffer from ascertainment bias.

Methods

Sampling and molecular data

A total of 1,593 individuals were genotyped for 135 AFLP markers, representing 13 zebu and 58 taurine breeds. The presence (genotype '1') or absence (genotype '0') of a band was scored considering AFLP as dominant markers, and occasional faint bands were considered as missing data. These samples were obtained from 23 countries from 9 distinct continental areas: Southern Asia (3 zebu breeds), Southwestern Asia (2 taurine breeds), Eastern Europe (3 taurine breeds), Central Europe (24 taurine breeds), Northern Europe (10 taurine breeds), Southern Europe (10 taurine breeds), Western Europe (8 taurine breeds), Western Africa (7 zebu breeds and 1 taurine breed), and South America (3 zebu breeds). This dataset builds on the data reported by Negrini *et al.* [16] by inclusion of samples of 20 additional breeds (Table 1). Individuals or markers presenting 5% or more missing data were excluded from the study. Further details on the AFLP protocol and repeatability of the genotypes obtained can be found in Additional file 1.

Genetic distances and distance-based clustering

We used *AFLPsurv v1.0* [22] to calculate three different measures of pairwise genetic distances between

Table 1 Continental areas, countries and breeds of taurine and zebu cattle sampled

Continental area	Country	Breed	Zebu	Taurine	Code	n ^a	n QC ^b
Southern Asia	India	<u>Hariana</u>	x		HAR	4	4
		<u>Tharparkar</u>	x		THA	4	4
	Pakistan	<u>Sahiwal</u>	x		SHA	4	4
Southwestern Asia	Turkey	Anatolian Black		x	ANB	24	23
		Turkish Gray		x	TGS	24	23
Eastern Europe	Hungary	<u>Hungarian Gray</u>		x	HUG	22	19
	Croatia	Istrian		x	ISR	24	23
Central Europe	Poland	<u>Polish Red</u>		x	POR	23	21
	Belgium	<u>Belgian Blue</u>		x	BEB	27	24
	France	Blond d'Aquitaine		x	BLM	20	19
		<u>Bretonne Pie Noir</u>		x	BPN	22	19
		<u>Charolais</u>		x	CHA	22	21
		<u>French Limousine</u>		x	LIM	25	21
		<u>Jersey</u>		x	JER	18	14
		Maine-Anjou		x	MAI	20	18
		<u>Montbéliard</u>		x	MON	22	22
		<u>Normande</u>		x	NOR	23	22
		Parthenaise		x	PAR	15	15
		Salers		x	SAL	20	20
		Switzerland		x	SWB	23	20
				x	ERI	19	19
				x	EVO	9	8
				x	SIM	21	19
	Italy	<u>Bruna</u>		x	BRU	33	29
		<u>Frisona</u>		x	FRI	47	44
		Grigio Alpina		x	GAL	21	19
		<u>Italian Limousine</u>		x	LMI	22	19
		<u>Piedmontese*</u>		x	PIM	22(21*)	21
		<u>Pezzata Rossa Italiana</u>		x	PRI	22	22
		<u>Rendena*</u>		x	REN	24(22*)	24
		<u>Valdostana Pezzata Rossa</u>		x	VPR	22	22
		Germany		x	GBP	20	20
				x	PIG	24	22
Northern Europe	England	Aberdeen Angus		x	ABA	20	15
		<u>Blacksided Trondheim</u>		x	BTR	22	21
	Norway	<u>Telemark</u>		x	TEL	22	22
		<u>Vestland Red Polled</u>		x	VPO	22	18
		Denmark		x	DAR	22	21
				x	JUT	22	18
	Finland	<u>Eastern Finn Cattle</u>		x	EFC	22	21
		<u>Finnish Ayrshire</u>		x	FAY	22	20
	Iceland	<u>Iceland Cattle</u>		x	ICE	22	22
	Sweden	<u>Swedish Red Polled</u>		x	SRP	22	20
Southern Europe	Italy	<u>Cabannina*</u>		x	CAB	22(20*)	20

Table 1 Continental areas, countries and breeds of taurine and zebu cattle sampled (Continued)

Western Europe	Spain	<u>Calvana</u>	x	CAL	40	38
		<u>Chianina*</u>	x	CHI	22(21*)	20
		<u>Cinisara</u>	x	CIN	8	7
		<u>Marchigiana</u>	x	MCG	22	20
		<u>Maremmiana</u>	x	MAR	45	45
		<u>Modicana</u>	x	MOD	12	12
		<u>Mucca Pisana</u>	x	MUP	40	39
		<u>Podolica*</u>	x	POD	22(22*)	20
		<u>Romagnola</u>	x	ROM	20	19
		<u>Asturiana de los Valles</u>	x	RAV	20	19
		<u>Betizu</u>	x	BET	20	18
		<u>DiLidia</u>	x	DLD	20	19
		<u>Menorquina</u>	x	MEN	20	19
		<u>Rubia Gallega</u>	x	RUG	20	20
		<u>Sayaguesa</u>	x	SAY	20	19
Western Africa	Cameroon	<u>Tudanca</u>	x	TUD	20	18
		Banyo Gudali	x	CBG	26	18
		Cameronian Red Bororo	x	CRB	25	20
		Cameronian White Fulani	x	CWF	23	23
	Guinea-Bissau	Ngaoundere Gudali	x	CNG	25	18
		<u>Guinean N'Dama</u>	x	GND	20	19
		Nigeria	Red Bororo	x	NRB	25
	Sokoto Gudali		x	NSG	25	25
	White Fulani		x	NWF	25	24
	South America	Brazil	Guzerat	x	GUZ	32
Nellore			x	NEL	32	21
Tabapuã			x	TAB	32	32
Total					1,593	1,470

^an: Number of samples before quality control.

^bn QC: Number of samples after quality control.

Quality control was performed by removing samples with 5% or more missing data.

*Breed/number of individuals used to test the repeatability of AFLP fingerprinting (see Additional files 1, 2, 3, 4, 5, 6, 7, 8 and 9).

Underlined breed names correspond to the samples described previously by Negrini *et al.* [16].

populations: F_{ST} [23], Nei's D [24] and Reynolds' distance [25]. We grouped animals according to breed or continental area. The three Southern Asian breeds were excluded in the analyses for individual breeds because of their low sample size ($n=12$). We used the *base* package in *R* v2.15.0 [26] to perform spectral decompositions on the matrices of pairwise genetic distances between groups in order to construct low-dimensional representations of the genetic relationships among the surveyed populations. The dissimilarities between pairs of groups were captured in $n-1$ dimensional spaces of n observations (eigenvectors), where n is the number of groups, via classical multi-dimensional scaling (CMDS) [27]. The proportion of genetic variance explained by each eigenvector was calculated by dividing its respective eigenvalue by the sum of all

eigenvalues, and expressed as percentages. Additionally, we applied the Neighbor-Net method to the distance matrices by using *SPLITSTREE v4.13.1* [28].

Expected heterozygosity and ancestry informative markers

With the particular interest of identifying geographical patterns in the extent of genetic diversity in the cattle breeds analyzed, we used *AFLPsurv v1.0* [22] to calculate expected heterozygosities for each continental area under the assumption of Hardy-Weinberg equilibrium. Essentially, the same values were obtained averaging per area over the expected heterozygosities of the separate breeds (data not shown). Additionally, we applied an *ad hoc* statistic to identify taurine and zebu ancestry

informative markers (i.e., AFLP markers with large differences in band presence frequencies between taurine and zebu breeds). For each AFLP marker, we computed the band presence frequency across all breeds, and then calculated the mean for the pool of taurine and zebu breeds. We then calculated the difference in band presence frequency as $\Delta f = f_{\text{taurine}} - f_{\text{zebu}}$. Positive and negative values indicate markers that are informative of taurine or zebu ancestry, respectively. We used thresholds of +0.55 and -0.55 to identify taurine and zebu ancestry informative AFLP markers, respectively. Finally, the average of band presence frequency of informative markers was computed for each breed in order to assess the relative level of taurine/zebu introgression across the investigated breeds.

Model-based clustering

We estimated individual ancestry coefficients as parameters of a statistical model, following the Bayesian approach implemented in *STRUCTURE* v2.2 [29]. This is referred as the admixture model adapted for AFLP markers with independent allele frequencies (see [29,30] for details). Briefly, it is assumed that the genomes of the sampled individuals derive from one or more of K ancestral populations, and the proportion of the individuals' ancestry from each one of these populations is estimated via a Markov chain Monte Carlo algorithm. The assumption that the alleles are independent (i.e. linkage equilibrium) is reasonable in the present study, as the AFLP panel used is sparse and the markers are unlikely to be closely located on the genome. We applied this model from $K = 1$ to $K = 60$, and ran 5 replicates of 150,000 iterations for each analysis after a burn-in of 100,000 iterations [31].

We applied two methods to identify the most likely number of ancestral populations underlying the observed data. The first method uses the ΔK statistic described by Evanno *et al.* [32], which is based on the rate of change in the log-likelihood of data between successive K values. The second method was abstracted from the approaches for model selection reviewed by Johnson & Omland [33], and is based on the concept of relative likelihood. First, the Akaike Information Criterion (AIC) is calculated for each model, from $K = 1$ to $K = 60$, as follows: $AIC = 2p - 2 \ln(L)$, where p and $\ln(L)$ are the number of parameters and the log-likelihood of the estimated model, respectively. Next, AIC differences are calculated for each model i as $\Delta_i = AIC_i - AIC_{\min}$, where AIC_{\min} corresponds to the lowest AIC among all models; and relative likelihoods are computed as $L_i = e^{-1/2\Delta_i}$. Then, relative likelihood values are normalized across all K models to produce Akaike weights $\omega_i = L_i / \sum_{j=1}^K L_j$. These can be interpreted as the probability that the

respective model is the one that presents the minimum information loss among all competing models, and was used as an alternative approach to estimate the optimal number of K .

Results

Quality control

After the exclusion of individuals exhibiting 5% or more missing genotypes, 1,470 animals remained from the initial set of 1,593 (see Table 1 for details). From a total of 135 genotyped AFLP loci, 8 were excluded due to missing data (>5%), and the final set of AFLP markers included 127 loci. As most of the analyses reported hereafter assume marker neutrality, the impact of the inclusion of putative markers under selection in all downstream analyses was evaluated. In all cases, the exclusion of candidate outlier markers resulted in no significant difference in the estimates of genetic distances and ancestry coefficients (Additional file 2). Therefore, all subsequent analyses were conducted using the entire set of 127 markers.

Genetic distance-based clustering

Different genetic distances were highly correlated (data not shown) and yielded consistent results (Additional file 3; Additional file 4: Figure S1; Additional file 5: Figure S2; Additional file 6: Figure S3; Additional file 7: Figure S4). We present the results obtained from Reynolds' distance (Figure 1), which was shown to be insensitive to variation in the number of markers [34].

The Nigerian zebu breeds Sokoto Gudali and White Fulani were the closest related populations (Reynolds' distance = 0.005). In contrast, in spite of a possible contribution of Spanish ancestry to Brazilian cattle, Brazilian and Spanish breeds are well separated with the largest distance between Nellore and the inbred Betizu (Reynolds' distance = 0.656).

The first two eigenvectors of the CMDS analysis of continental groups of cattle (Figure 1B) explained together 79.4% of the total genetic variance, and were centered on Southwestern Asian taurine cattle. The first eigenvector corresponds to the difference between taurine and zebu cattle with Southern Asian and South American zebu clustered together, and an intermediate position of Western African zebu cattle. The second eigenvector adds a geographical component correlating with the latitude of the region of origin of cattle populations (Figure 1A-B). The Neighbor-Net clustering method produced results similar to those found in the CMDS analysis (Figure 1C).

Model-based clustering

The log-likelihoods obtained from the admixture model with independent allele frequencies, assuming $K = 1$ to

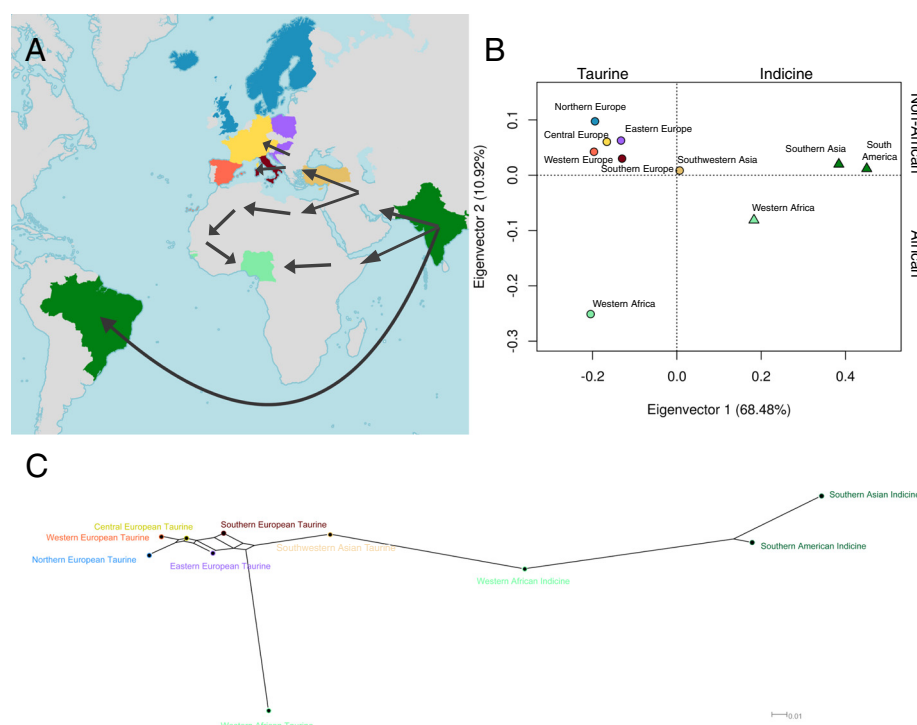


Figure 1 Reynolds' distance-based clustering of cattle according to continental areas. **A)** Continental areas sampled. Light brown = Southwestern Asia, purple = Eastern Europe, yellow = Central Europe, dark blue = Northern Europe, dark red = Southern Europe, orange = Western Europe, light green = Western Africa, dark green = Southern Asia and South America. Arrows indicate cattle migration routes. **B)** Classical multi-dimensional scaling plot. Circles: taurine cattle; triangles: zebu cattle. Percentages inside brackets correspond to the variance explained by each respective eigenvector. **C)** Neighbor-Net clustering. Nodes represent continental areas and edges are proportional to genetic distances.

$K = 60$, were compared using ΔK and AIC weights in order to identify the most likely number of ancestral populations underlying the samples. Both ΔK and AIC weights selected the model with $K = 2$ as the most likely among all competing models (Additional file 8: Figure S5). Assuming the two inferred clusters approximate the founder *B. taurus* and *B. indicus* populations (Figure 2A), we found variable levels of zebu introgression across taurine cattle breeds from all continental areas, which were especially marked in Southwestern Asian taurines. While South American zebu breeds did not present evident taurine introgression, this was detected in all Western African zebu breeds.

Higher K values were not supported by both ΔK and AIC weights and were not in agreement with genetic distances (data not shown). This indicates that models with $K > 2$ were susceptible to stochastic errors and represented poorly the underlying ancestry components of our samples. This may be due to model overfitting, by estimation of more parameters than allowed by the observed data. Hence, for our dataset, the model-based clustering analysis was limited to $K = 2$ due to the low number of dominant markers and estimation of unobserved genotypes.

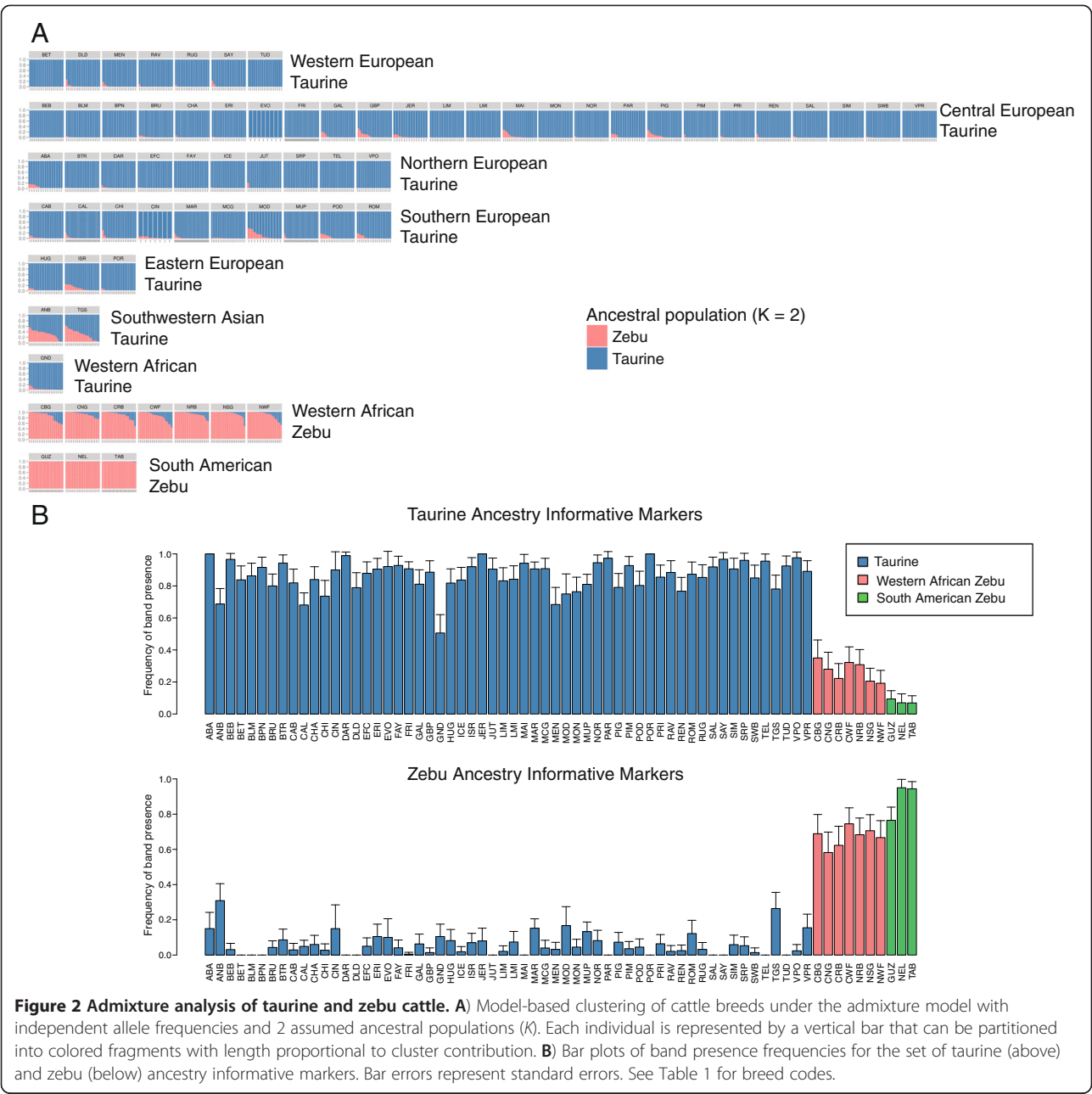
Ancestry informative markers and expected heterozygosities

We identified 6 taurine and 5 zebu ancestry informative markers via Δf , and calculated the average band presence frequency for these markers across all breeds (Figure 2B). We observed that the taurine markers had in Western African zebus a higher frequency of band presence than in South American zebu, and the opposite was also found for zebu markers.

We found a geographical pattern of decrease in the expected heterozygosity in taurine cattle, declining from Southwestern Asia to Western Europe and Western Africa (Additional file 9: Figure S6). Despite the limited sample size, Southern Asian zebus were estimated to be more diverse than the pools of taurine breeds. The estimate obtained for the closely related South American zebu was slightly lower than in Southwestern Asian taurines, but still higher than in European cattle. Furthermore, Southern Asian and Western African zebu exhibited the highest expected heterozygosity among all continental groups analyzed.

Discussion

The performance of AFLP technology in cattle was previously assessed and reported to produce genotyping data



with an error rate equal to or less than 2% across laboratories [16], which is consistent with the repeatability of the data reported in the present study (Additional file 1). Here, we revisited the use of AFLPs to investigate the relationship among 13 zebu and 53 taurine cattle breeds. As AFLP markers are discovered as samples are genotyped, the assessment of genetic structure and differentiation reported in this article is free of ascertainment bias.

As expected, the largest genetic distances were found between zebu and taurine breeds (Additional file 3). The Bayesian-clustering analysis also highlighted that these populations descend from distinct genetic pools

(Figure 2). We found a decrease of the genetic diversity correlating with geographical distance to Southwestern Asia (Additional file 9: Figure S6). This observation is in agreement with the mitochondrial DNA (mtDNA) findings of Troy *et al.* [35], which suggested a Southwestern Asian origin of European cattle with Anatolia or the Fertile Crescent as the most likely centre of taurine cattle domestication. Hence, the loss of diversity with increasing distance from the most plausible domestication centre as observed here is in line with the hypothesis that the ancestral taurine genetic pool was derived from the wild aurochs captured in Southwestern Asia. Apparently,

any introgression from European or African aurochs was not at such a large scale that it effectively counteracted the loss of diversity during migration from Southwestern Asia.

Using sequence data of 17 genes, spanning 37 kb, Murray *et al.* [36] found the nucleotide and haplotype diversity in *B. indicus* to be higher than in *B. taurus*. In the present study, we also found that the expected heterozygosity in the South American zebu breeds was higher than in the European taurine breeds. Considering that the South American zebu breeds analyzed here were introduced in the American continent in the early 20th century by import of Indian animals, this finding is also consistent with a separate origin of *B. indicus* in South Asia.

The expected heterozygosity in Southern Asian cattle was estimated to be higher than the closely related South American breeds. Although this finding is consistent with loss of diversity during sampling and importation of animals to South America, Southern Asian cattle were represented by few samples in our dataset, and the assessment of the extent of genetic diversity in this continental group is limited. However, these results support that the *B. indicus* species are at least as diverse as *B. taurus* cattle.

The CMDS and Neighbor-Net analyses showed that zebu cattle from South America are more closely related to Southern Asian cattle than Western African zebu (Figure 1). Furthermore, except for Southern Asian zebu, Western African zebu breeds presented the highest expected heterozygosity among all continental groups. Most likely, this was due to a relatively higher level of admixture [5,37,38].

The closer proximity of Western African zebu to taurine cattle in the CMDS plot and in the Neighbor-Net of Reynolds' distances also suggests that African zebras are more admixed with taurine cattle than South American zebras (Figure 1). This observation is reinforced by the model-based clustering and the ancestry informative markers analyses, where these African breeds seemed to carry substantial levels of taurine introgression (Figure 2). This may reflect that zebu cattle and taurine-zebu cross-breeds in Africa resulted from crosses between taurine dams and zebu sires as shown by their taurine mtDNA haplotypes: import of zebu sires started in the 2nd millennium BC and was stimulated by the Arabian invasions in the 7th century [4,39]. However, it is also plausible that this taurine inheritance played a role in local adaptation. For instance, trypanosomiasis is endemic in the Western Sub-Saharan region, and whereas indigenous taurines are tolerant, zebu may show variable susceptibility.

Similar crossbreeding was carried out in South America. When in the early 20th century the import of large numbers of zebu cattle to Brazil started, the indigenous herds mainly consisted of descendants from the taurine cattle imported since the late 15th century after the

discovery of America. The model-based clustering analysis clearly showed a genetic composition of Brazilian zebu close to their Indian ancestors (Figure 2A-B), indicating intensive backcrossing to zebu bulls during several generations. So while mtDNA is a fingerprint of the historical origin of the herd and is probably randomly segregating [40,41], the nuclear genome has been subject to directional selection against taurine haplotypes via backcrossing. Thus, artificial selection may have retained taurine haplotypes only if these were linked to favourable traits (e.g., weight, carcass, etc.). Applying whole genome sequence data or a high density SNP array may be useful to identify taurine haplotypes favoured by selection in these populations.

Ancestry informative markers also detected zebu introgression in the taurine gene pool (Figure 2). The highest level of introgression was found in Southwestern Asia, as previously observed with microsatellites [37]. This event likely contributed to the highest diversity that is observed in this area and, therefore, should not be attributed entirely to the vicinity of Southwestern Asian breeds to the putative *B. taurus* centre of domestication. A low level of admixture was also detected in Southern and Central Italian breeds, the Sicilian Cinisara and Modicana in particular, confirming a previous report [42]. The zebu admixture appears to decrease across the Alps towards Central and Western Europe with few exceptions (e.g., Aberdeen Angus). Interestingly, we confirmed the low level of *B. indicus* introgression in Pinzgauer breed postulated by Caroli *et al.* [43] on the basis of casein haplotype structure in Austria, but did not detect substantial zebu ancestry in the Piedmontese breed as previously suggested [44]. Given the limited number of ancestry informative markers (5 zebu and 6 taurine), these results are only indicative and can be confounded by stochastic variation.

Conclusions

We used AFLP markers to set an unbiased baseline for multi-breed taurine and zebu cattle genetic structure and divergence. These markers suggested that zebu breeds are at least as diverse as taurine cattle, but further investigation is needed to determine if zebu cattle is more diverse than taurine cattle. We found a gradual loss of diversity in taurine breeds departing from the domestication centre, which is consistent with previous findings. Western African zebu breeds are more genetically distant to Indian zebras than South American zebu cattle by substantial taurine introgression. Although the South American zebras also have maternal taurine introgression, most of the taurine component of the nuclear genome seems to have disappeared through backcrossing. Furthermore, the AFLP data indicated limited zebu introgression in the Italian Podolian breeds.

Additional files

Additional file 1: AFLP protocol and repeatability.

Additional file 2: Bias and inflation of non-neutral markers.

Additional file 3: Matrices of pairwise genetic distances between cattle breeds and continental areas. BIND = *B. indicus*, BTAU = *B. taurus*, SWA = Southwestern Asia, EE = Eastern Europe, CE = Central Europe, NE = Northern Europe, SE = Southern Europe, AF = Africa, AS = Asia, SA = South America. See Table 1 for breed codes.

Additional file 4: Figure S1. Classical multi-dimensional scaling analysis between continental areas using three different measures of genetic distance. Percentages inside brackets correspond to the variance explained by the eigenvector. Abbreviations as for Additional file 1.

Additional file 5: Figure S2. Classical multi-dimensional scaling analysis between cattle breeds using three different measures of genetic distance. Percentages inside brackets correspond to the variance explained by each respective eigenvector. See Table 1 for breed codes.

Additional file 6: Figure S3. Neighbor-net clustering of cattle breeds according to continental area using three different measures of genetic distance. Abbreviations as for Additional file 1.

Additional file 7: Figure S4. Neighbor-net clustering of cattle breeds using three different measures of genetic distance. See Table 1 for breed codes.

Additional file 8: Figure S5. Model selection for the most probable number of ancestral populations according to two criteria (see Methods).

Additional file 9: Figure S6. Bar plot of expected heterozygosities for each continental area. Error bars represent standard errors. Abbreviations as for Additional file 1.

Abbreviations

SNP: Single nucleotide polymorphism; 50 k: Illumina® BovineSNP50 BeadChip assay; AFLP: Amplified fragment length polymorphism; CMDs: Classical multi-dimensional scaling; AIC: AKAIKE information criterion; mtDNA: Mitochondrial DNA.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PAM and JFG conceived and led the coordination of the study. The European Cattle Genetic Diversity Consortium contributed with data. LC, RN, JAL and GE contributed to the study design. YTU, LB, GL and PAM performed data analyses. YTU drafted the manuscript. PAM, JFG, YTU, LB, JAL, LC, RN and GE interpreted the results and contributed to edit the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We wish to thank Dr. Elia Vajana for the help provided in analyzing the data on the consistency of AFLP profiles. We are also grateful to three anonymous reviewers, whose suggestions have substantially improved the quality of the manuscript. This research was supported by: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) - process 2011/16643-2 and 2013/12829-0. Mention of trade name proprietary product or specified equipment in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the authors or their respective institutions. The following members of the European Cattle Genetic Diversity Consortium contributed to this study: France: K. Moazami-Goudarzi, INRA, Jouy-en-Josas; UK: J. Williams and P. Wiener, Roslin Institute; Norway: I. Olsaker, Norwegian School of Veterinary Science, Oslo; Finland: J. Kantanen, Agrifood Research Finland (MTT), Jokioinen; Spain: S. Dunner and J. Cañón, Universidad Complutense de Madrid; C. Rodellar, I. Martín-Burriel, Veterinary Faculty, Zaragoza; Italy: A. Valentini, Università della Tuscia, Viterbo; M. Zanotti, Università degli Studi di Milano; Denmark: L.-E. Holm, Aarhus University, Tjele; Iceland: E. Eythorsdottir, Agricultural Research Institute, Reykjavik; Belgium: G. Mommens, Dr. Van Haeringen Polygen, Malle; The Netherlands: I.J. Nijman, Utrecht University; Switzerland: G. Dolf, University of Berne; Ireland: D.G. Bradley, Trinity College, Dublin.

Author details

¹Faculdade de Ciências Agrárias e Veterinárias, UNESP - Univ Estadual Paulista, Jaboticabal, São Paulo 14884-900, Brazil. ²Institute of Zootechnics, Università Cattolica del Sacro Cuore, Piacenza, Italy. ³BioDNA Biodiversity and Ancient DNA Research Centre, Università Cattolica del Sacro Cuore, Piacenza, Italy. ⁴Faculty of Veterinary Medicine, Utrecht University, Utrecht, Netherlands. ⁵Institute of Animal Breeding and Genetics, Justus-Liebig University, Giessen 21b, 35390, Ludwigstrasse, Germany. ⁶Faculdade de Medicina Veterinária de Araçatuba, UNESP - Univ Estadual Paulista, Araçatuba, São Paulo 16050-680, Brazil.

Received: 14 October 2013 Accepted: 9 April 2014

Published: 17 April 2014

References

- Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E, Weigend S, GLOBALDIV Consortium: Genetic diversity in farm animals - a review. *Anim Genet* 2010, **41**:6-31.
- Bruford MW, Bradley DG, Luikart G: Genetic analysis reveals complexity of livestock domestication. *Nat Rev Genet* 2003, **4**:900-910.
- Loftus RT, MacHugh DE, Bradley DG, Sharp PM, Cunningham P: Evidence for two independent domestications of cattle. *Proc Natl Acad Sci U S A* 1994, **91**:2757-2761.
- Ajmone-Marsan P, Garcia JF, Lenstra JA, Globaldiv Consortium: On the origin of cattle: how aurochs became cattle and colonized the world. *Evol Anthropol* 2010, **19**:148-157.
- Hanotte O, Bradley DG, Ochieng JW, Verjee Y, Hill EW, Rege JEO: African pastoralism: genetic imprints of origins and migrations. *Science* 2002, **296**:336-339.
- Bovine HapMap C, Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Sonstegard TS, Williams JL, Diallo B, Hailemariam L, Martinez ML, Morris CA: Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* 2009, **324**(5926):528-532.
- Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilki J, Seabury CM, Caetano AR, Johnson GS, Breneman RA, Hanotte O, Eggert LS, Wiener P, Kim JJ, Kim KS, Sonstegard TS, Van Tassell CP, Neibergs HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD, Taylor JF: Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc Natl Acad Sci* 2009, **106**:18644-18649.
- Gautier M, Laloë D, Moazami-Goudarzi K: Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS One* 2010, **5**:e13038.
- McTavish EJ, Decker JE, Schnabel RD, Taylor JF, Hillis DM: New World cattle show ancestry from multiple independent domestication events. *Proc Natl Acad Sci* 2013, **110**:E1398-406.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP: Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One* 2009, **4**:e5350.
- Vos P, Hogers R, Bleeker M, Reijers M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M: AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res* 1995, **23**:4407-4414.
- Bensch S, Akesson M: Ten years of AFLP in ecology and evolution: why so few animals? *Mol Ecol* 2005, **14**:2899-2914.
- Ajmone-Marsan P, Valentini A, Cassandro M, Vecchiotti-Antaldi G, Bertoni G, Kuiper M: AFLP markers for DNA fingerprinting in cattle. *Anim Genet* 1997, **28**:418-426.
- Ajmone-Marsan P, Otsen M, Valentini A, Bertoni G, Kuiper MTR, Lenstra JA: Genetic distances within and across cattle breeds as indicated by biallelic AFLP markers. *Anim Genet* 2002, **33**:280-286.
- Savelkoul PHM, Aarts HJM, Dijkshoorn L, Duims B, De Haas J, Otsen M, Schouls L, Lenstra JA: Amplified fragment length polymorphism (AFLP™), the state of an art. *J Clin Microbiol* 1999, **37**:3083-3091.
- Negrini R, Nijman IJ, Milanese E, Moazami-Goudarzi K, Williams JL, Erhardt G, Dunner S, Rodellar C, Valentini A, Bradley DG, Ajmone Marsan P, Lenstra JA, The European Cattle Genetic Diversity Consortium: Differentiation of European cattle by AFLP fingerprinting. *Anim Genet* 2007, **38**:60-66.

17. Ollivier L, Alderson L, Gandini G, Foulley JL, Haley CS, Joosten R, Rattink AP, Harlizius B, Groenen MAM, Amigues Y, Boscher MY, Russel G, Law A, Davoli R, Russo V, Matassino D, Desautels C, Firmland E, Bagga M, Delgado JV, Vega-Pla JL, Martinez AM, Ramos AM, Glodek P, Meyer JM, Plastow GS, Siggins KW, Archibald AL, Milan D, San Cristobal M, et al: **An assessment of European pig diversity using molecular markers: partitioning of diversity among breeds.** *Conserv Genet* 2005, **6**:729–741.
18. SanCristobal M, Chevalet C, Peleman J, Heuven H, Brugmans B, Van Schriek M, Joosten R, Rattink AP, Harlizius B, Groenen MA, Amigues Y, Boscher MY, Russell G, Law A, Davoli R, Russo V, Desautels C, Alderson L, Firmland E, Bagga M, Delgado JV, Vega-Pla JL, Martinez AM, Ramos M, Glodek P, Meyer JN, Gandini G, Matassino D, Siggins K, Laval G, et al: **Genetic diversity in European pigs utilizing amplified fragment length polymorphism markers.** *Anim Genet* 2006, **37**:232–238.
19. Ritz LR, Glowatzki-Mullis ML, MacHugh DE, Gaillard C: **Phylogenetic analysis of the tribe Bovini using microsatellites.** *Anim Genet* 2000, **31**:178–185.
20. Hassani A, Ropiquet A: **Molecular phylogeny of the tribe Bovini (Bovidae, Bovinae) and the taxonomic status of the Koupriy, Bos sauveli Urbain 1937.** *Mol Phylogenet Evol* 2004, **33**:896–907.
21. Buntjer JB, Otsen M, Nijman IJ, Kuiper MT, Lenstra JA: **Phylogeny of bovine species based on AFLP fingerprinting.** *Heredity* 2002, **88**:46–51.
22. Vekemans X, Beauwens T, Lemaire M, Roldan-Ruiz I: **Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size.** *Mol Ecol* 2002, **11**:139–151.
23. Wright S: **The genetical structure of populations.** *Ann Eugen* 1951, **15**:323–354.
24. Lynch M, Milligan BG: **Analysis of population genetic structure with RAPD markers.** *Mol Ecol* 1994, **3**:91–99.
25. Reynolds J, Weir BS, Cockerham CC: **Estimation of the coancestry coefficient: basis for a short-term genetic distance.** *Genetics* 1983, **105**:767–779.
26. R Development Core Team: **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing; 2008 [http://www.R-project.org]
27. Mardia KV: **Some properties of classical multi-dimensional scaling.** *Commun on Stat-Theory and Methods* 1978, **A7**:1233–1241.
28. Bryant D, Moulton V, Neighbor-Net: **An Agglomerative method for the construction of phylogenetic networks.** *Mol Biol Evol* 2004, **21**:255–265.
29. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945–959.
30. Falush D, Stephens M, Pritchard JK: **Inference of population structure using multilocus genotype data: dominant markers and null alleles.** *Mol Ecol Notes* 2007, **7**:574–578.
31. Gilbert KJ, Andrew RL, Bock DG, Franklin MT, Kane NC, Moore JS, Moyers BT, Renaut S, Rennison DJ, Veen T, Vines TH: **Recommendations for utilizing and reporting population genetic analyses: the reproducibility of genetic clustering using the program structure.** *Mol Ecol* 2012, **21**:4925–4930.
32. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.** *Mol Ecol* 2005, **14**:2611–2620.
33. Johnson JB, Omland KS: **Model selection in ecology and evolution.** *Ecol Evol* 2004, **19**:101–108.
34. Lenstra JA, Groeneveld LF, Eding H, Kantanen J, Williams JL, Taberlet P, Nicolazzi EL, Sölkner J, Simianer H, Ciani E, Garcia JF, Bruford MW, Ajmone-Marsan P, Weigend S: **Molecular tools and analytical approaches for the characterization of farm animal genetic diversity.** *Anim Genet* 2012, **43**:483–502.
35. Troy CS, MacHugh DE, Bailey JF, Magee DA, Loftus RT, Cunningham P, Chamberlain AT, Sykes BC, Bradley DG: **Genetic evidence for near-eastern origins of European cattle.** *Nature* 2001, **411**:1088–1091.
36. Murray C, Huerta-Sanchez E, Casey F, Bradley DG: **Cattle demographic history modelled from autosomal sequence variation.** *Philos Trans R Soc Lond B Biol Sci* 2010, **365**(1552):2531–2539.
37. Freeman AR, Bradley DG, Nagda S, Gibson JP, Hanotte O: **Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle.** *Anim Genet* 2006, **37**:1–9.
38. Freeman AR, Hoggart CJ, Hanotte O, Bradley DG: **Assessing the relative ages of admixture in the bovine hybrid zones of Africa and the near east using X chromosome haplotype mosaicism.** *Genetics* 2006, **173**:1503–1510.
39. Bradley DG, MacHugh DE, Cunningham P, Loftus RT: **Mitochondrial diversity and the origins of African and European cattle.** *Proc Natl Acad Sci U S A* 1996, **93**:5131–5135.
40. Meirelles FV, Rosa AJM, Lobo RB, Garcia JM, Smith LC, Duarte FAM: **Is the American zebu really *Bos indicus*?** *Genet Mol Biol* 1999, **22**:543–546.
41. Paneto JC, Ferraz JB, Balieiro JC, Bittar JF, Ferreira MB, Leite MB, Merighe GK, Meirelles FV: ***Bos indicus* or *Bos taurus* mitochondrial DNA: comparison of productive and reproductive breeding values in a Guzerat dairy herd.** *Genet Mol Res* 2008, **7**:592–602.
42. Pieragostini E, Scaloni A, Rullo R, Di Luccia A: **Identical marker alleles in podolic cattle (*Bos Taurus*) and Indian zebu (*Bos indicus*).** *Comp Biochem Physiol B Biochem Mol Biol* 2000, **127**:1–9.
43. Caroli A, Rizzi R, Lühken G, Erhardt G: **Short communication: milk protein genetic variation and casein haplotype structure in the original pinzgauer cattle.** *J Dairy Sci* 2009, **93**:1260–1265.
44. Merlin P, Di Stasio L: **Study on milk proteins loci in some decreasing Italian cattle breeds.** *Ann Genet Sel Anim* 1982, **14**:17–28.

doi:10.1186/1471-2156-15-47

Cite this article as: Utsunomiya et al.: Revisiting AFLP fingerprinting for an unbiased assessment of genetic structure and differentiation of taurine and zebu cattle. *BMC Genetics* 2014 **15**:47.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

